

Labs for course #412

Analyzing Microarray Data using the mAdb System

February 16-17, 2005 1:00pm- 4:00pm

- First, look at the questions on the bottom of each page. Write down the answers while going through the steps on the page.
- Keep the browser NOT maximized so multiple windows can be distinguished.

Lab 1. Copying a Training Dataset

Goal: To copy a dataset into user's temporary area and to inspect dataset features.

4

[Copy](#) Small, Round Blue Cell Tumors (SRBCTs) data containing 88 Arrays with 2308 Features to your temporary area.

Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Khan J, Wei JS, Ringer M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer PS, Nature Medicine Vol 7, Num 6, 601-673 (2001)

[Copy](#) Subset of NEJM data containing 60 Arrays with 1626 Features to your temporary area. Includes Feature Property Filters.


1. Open a web browser and type the URL for the mAdb home page, **for training class: <http://madb-training.cit.nih.gov>** - use login on name tent and password on board. Others can use <http://madb.nci.nih.gov> (NIAID users <http://madb.niaid.nih.gov>) and log in with your mAdb account.

2. Click the first bullet, **mAdb Gateway** to access mAdb Gateway Web page

3. On the mAdb Gateway Web page, Click the link **Access Training/Public Dataset** on the bottom of the page. A page for copying three training datasets will be presented.

4. You can choose to work with either "Small, Round Blue Cell Tumors (SRBCT) dataset" or "NEJM Dataset". Click link **Copy** to copy the dataset into your temporary area.

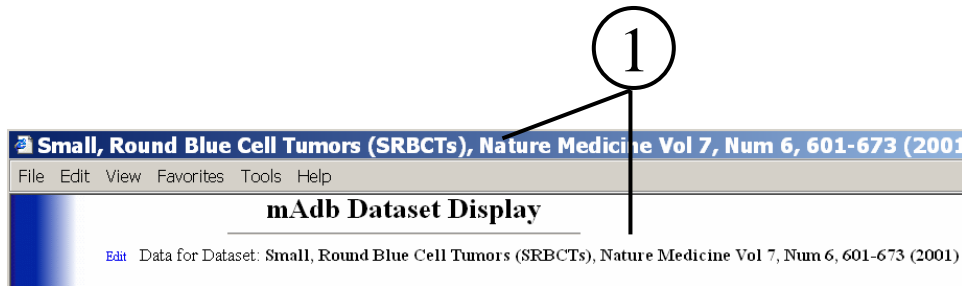
5. After copying the data, you will see the temporary dataset area. Click link **Open** on the selected dataset line. A mAdb Dataset Display page will be displayed.

Temporary Datasets		Created	Containing		Need Help? 
			Arrays	Genes	
Edit	Small, Round Blue Cell Tumors (SRBCTs), Nature Medicine Vol ...	Aug 26 6:00:00pm	88	2308	Open Refresh
Edit	NEJM - 3 Classes	Aug 26 5:23:18pm	60	1629	Open Refresh

5

Questions:

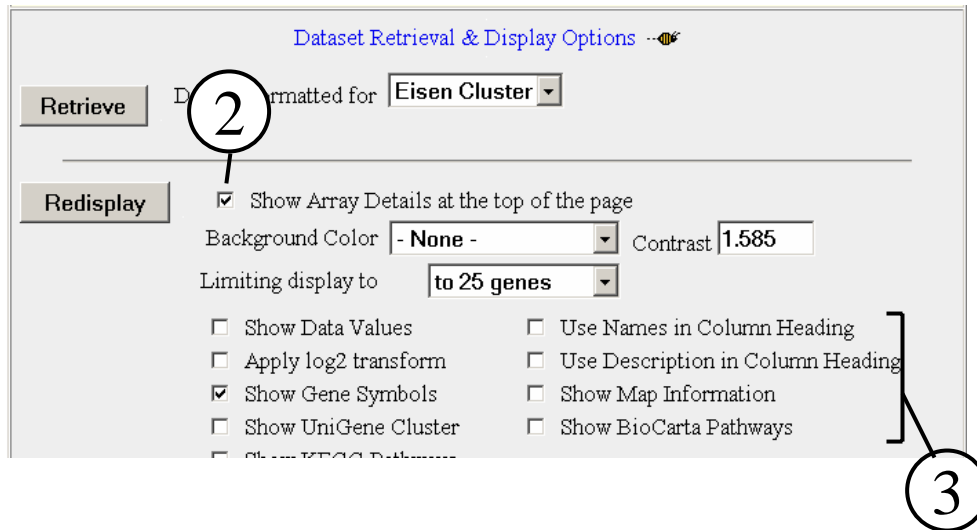
1. How many genes and how many arrays do you have in your dataset?



1. On the mAdb Dataset Display page, review the title bar, or the dataset description on top of the page. This tells you which dataset you are displaying.

2. In the **Dataset Retrieval & Display Options** panel, check the **Show Array Details at the top of the page** option. Then click **Redisplay** button. The names and short descriptions of arrays in the dataset will be displayed on the top of the page. Look for naming conventions of the array and then answer the question below. This information will be used in the next lab.

After reviewing the array details, it is recommended to uncheck the **Show Array Details at the top of the page** option. Click **Redisplay** to hide the array details on the top of the page.



3. Check or uncheck other display options of interest, and click **Redisplay** button to display or hide the relevant information. Uncheck **Show Data Values** and set **Background Color** to None will make it easier to view other annotations.

Questions:

1. How many experiment groups can you identify in this dataset by their naming conventions? Write down the naming conventions for each group.

Lab 2. Assigning Group Labels

Goal: To partition arrays into groups according to experiment design by assigning group labels.

Filtering/Grouping/Analysis Tools

Choose a Tool **Filter/Group by Array Properties** and **Proceed**

Interactive Graphical Viewers

Choose a Viewer **MDS: MultiDimensional Scaling** and **View**

1. In the Filtering/Grouping/Analysis section, choose the **Filter/Group by Array Properties Tool**

2. Click on **Proceed**

A new page will be displayed with options for assigning arrays into groups by the naming convention of **Array Name** or **Short Description**.

3. For the SRBC dataset, use EWS, BL, NB, RMS as matching patterns. Select **Array Name** and **Begins with** from the drop down list for each group. Samples with name beginning with "Test" are excluded from the grouped subset.

For the NEJM dataset, use GCB, ABC, and Type as matching patterns. Select **Array Name** and **Begins with** from the drop down list for each group.

4. The grouped results are stored as a new subset. Enter an appropriate label for this subset.

5. Click on **Submit**. There is no "Waiting" page, the new grouped subset will be directly displayed when the Group/Filtering process is completed.

Group A **Array Name** **Begins with** EWS

Group B **Array Name** **Begins with** BL

Group C **Array Name** **Begins with** NB

Group D **Array Name** **Begins with** RMS

Group E **Array Name** **Contains**

Expand the number of possible Group Designations to 10 , 15 , 20 or 26 groups.

Subset Label: **My Grouped Dataset**

Submit **Cancel**

mAdb Dataset Display

[Edit](#) Data for Subset: **My Grouped Dataset**

from Dataset: **Small, Round Blue Cell Tumors (SRBCTs)**, Nature Medicine Vol 7, Num 6, 601-673 (2001)

1

Filter/Group by Array Property

88 arrays and 2308 genes in the original dataset

63 arrays and 2308 genes in the output dataset.

Filter/Group by Array Property:

Group A: Array/Set Name Begins with 'ews'

Group B: Array/Set Name Begins with 'bl'

Group C: Array/Set Name Begins with 'nb'

Group D: Array/Set Name Begins with 'rms'

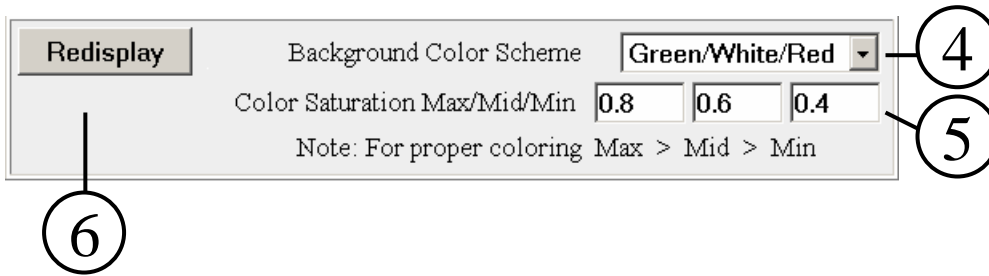
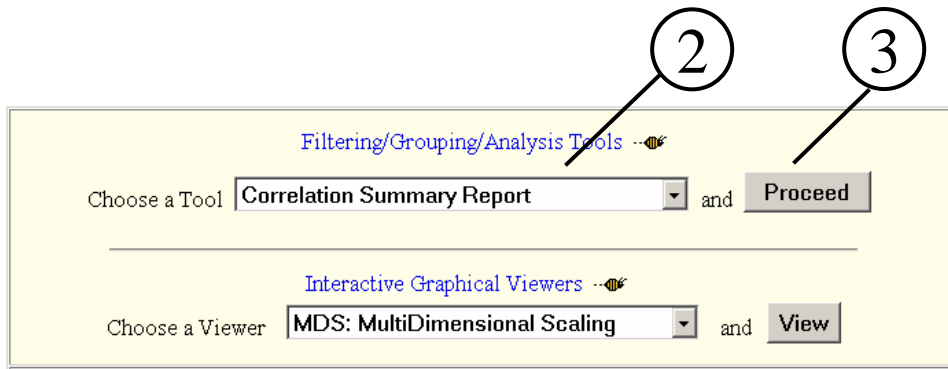
1. Examine the grouped subset through the dataset description and history on top of the Dataset Display page.

Questions:

1. How many arrays are filtered out in the grouped dataset?
2. What are they? Hint use “Array Order Designation/Filtering Tool”.
3. How many arrays do you have in each group? Write down the group designations for each tumor type.

Lab 3. Generating a Correlation Summary Report

Goal: To study the correlation of expression data among samples in the dataset.



1. Verify that the current dataset is My Grouped Dataset through title bar or dataset description. (See Lab 1, Dataset display section for details)

2. In the Filtering/Grouping/Analysis section, choose the **Correlation Summary Report** Tool. (You may have to scroll down the Tool dropdown list to find it on the bottom.)

3. Click on **Proceed**.

mAdb Correlation Report page will be displayed with a table of correlation results.

4. Change the **Background Color Scheme** to **Green/White/Red**.

5. Inspect the values of the correlation tables and set the values for **Color Saturation**. For SRBCT dataset, use 0.8, 0.6, 0.4. For NEJM-3 class dataset use 0.3, 0.0, -0.3.

6. Click on **Redisplay** button. Correlation table will be colored according to the correlations.

1. The image shows part of the correlation table. The color pattern uses green for good correlations and red for poor correlations.
2. Each correlation number represent a pair-wise correlation calculation between 2 samples. It can be clicked to display a scatter plot between the 2 samples. Click on a larger number to display the scatter plot for 2 correlated samples.
3. Click a small number to display a scatter plot for 2 poorly correlated samples .

② ③

B	B	B	B	B	B	B	B	C	C	C	C	C	C	C	C	C	C	C	C
#24	#25	#26	#27	#28	#29	#30	#31	#32	#33	#34	#35	#36	#37	#38	#39	#40	#41	#42	#43
#24 B	0.841	0.812	0.793	0.707	0.682	0.712	0.719	0.500	0.500	0.518	0.661	0.652	0.599	0.607	0.618	0.599	0.634	0.604	0.615
	#25 B	0.846	0.834	0.759	0.728	0.759	0.781	0.555	0.554	0.568	0.668	0.674	0.602	0.654	0.698	0.640	0.654	0.654	0.651
		#26 B	0.896	0.751	0.708	0.763	0.778	0.521	0.536	0.569	0.624	0.617	0.538	0.618	0.643	0.593	0.623	0.623	0.639
			#27 B	0.725	0.669	0.742	0.755	0.521	0.555	0.587	0.619	0.628	0.544	0.592	0.655	0.600	0.629	0.665	0.656
				#28 B	0.862	0.897	0.855	0.669	0.674	0.693	0.646	0.601	0.590	0.614	0.536	0.593	0.619	0.636	0.612
					#29 B	0.904	0.832	0.700	0.634	0.651	0.574	0.582	0.535	0.604	0.533	0.590	0.593	0.574	0.536
						#30 B	0.876	0.686	0.619	0.646	0.581	0.564	0.515	0.606	0.523	0.587	0.592	0.586	0.563
							#31 B	0.710	0.644	0.666	0.574	0.559	0.509	0.631	0.571	0.591	0.588	0.587	0.579
								#32 C	0.654	0.711	0.642	0.649	0.516	0.785	0.683	0.699	0.671	0.668	0.675
									#33 C	0.826	0.622	0.615	0.803	0.606	0.572	0.610	0.670	0.688	0.628
										#34 C	0.677	0.664	0.662	0.677	0.663	0.667	0.716	0.745	0.711
											#35 C	0.836	0.776	0.804	0.749	0.752	0.835	0.837	0.840
												#36 C	0.708	0.748	0.760	0.776	0.780	0.818	0.812
													#37 C	0.651	0.593	0.656	0.765	0.762	0.709
														#38 C	0.790	0.762	0.782	0.752	0.755
															#39 C	0.736	0.766	0.754	0.791
																#40 C	0.809	0.785	0.751
																	#41 C	0.873	0.777
																		#42 C	0.821
																			#43 C

Questions:

1. Describe the general color pattern of the correlation table. Are correlation numbers within a group better (more green) than between groups (more red)?
2. How is the scatter plot of a good correlation different from a plot of a poor correlation?

Lab 3. Generating a Correlation Summary Report

Lab 4. Filtering Data

Goal: To pre-process a dataset for further analysis by filtering out genes with low variance or with many missing values.

Filtering/Grouping/Analysis Tools

Choose a Tool **Additional Filtering Options** and **Proceed**

Interactive Graphical Viewers

Choose a Viewer **MDS: MultiDimensional Scaling** and **View**

1. Use back button on web browser to return to previous Dataset Display page. Verify that the current dataset is My Grouped Dataset. (See Lab 1, Dataset display section for details).

2. In the Filtering/Grouping/Analysis section, choose the **Additional Filtering Options** Tool.

3. Click on **Proceed**

Data Filtering Options page will be displayed with options for Missing Value Filters and Gene Filters. Be careful to check the "checkboxes" along putting in values in step 4-9.

4. Select the check box for **Genes: Require values in >=**

5. Set the value to **95% of arrays**.

6. Select the check box for **Variance (Gene Vector) percentile**

7. Set the value >= **90%**

8. The filtered results are stored as a new data subset. Enter an appropriate Label for this subset.

9. Click on **Filter**. Filtering will be performed and the results stored as a new subset. There is no "Waiting" page, the new subset will be directly displayed when the Filtering process is completed.

Missing Value Filters

☒ Genes: Require values in >= 95 % of Arrays

☐ Arrays: Require values in >= 80 % of Genes

Gene Filters

☐ Ratio >= 20 in >= 1 Arrays

☐ Apply Symmetrically

☐ Ratio >= 2 in >= 2 Arrays OR Ratio <= 0.5 in >= 2 Arrays

☐ Average Ratio >= 2

☐ Apply Symmetrically

☐ Max (Ratio) / Min (Ratio) >= 3

☒ Variance (Gene Vector) percentile >= 90 %

Subset Label: Gene>=95%, Variance >=90%

Filter Cancel

mAdb Dataset Display

[Edit](#) Data for Subset: **Gene** >=95%, **Variance** >=90%

from Dataset: **Small, Round Blue Cell Tumors (SRBCTs)**, Nature Medicine Vol 7, Num 6, 601-673 (2001)

The filter input data set contained 63 arrays and 2308 genes.
The filtered output data set contains 63 arrays and 230 genes.
No genes excluded for being present in less than 95% (60) arrays.
2078 genes excluded where variance is in the lowest 90 percentile (Variance<1.60).

View the complete [History](#).

This dataset was constructed from the supplemental data posted at

Thu Oct 9 17:57:21 EDT 2003

Filter/Group by Array Property
88 arrays and 2308 genes in the [original dataset](#) dataset
63 arrays and 2308 genes in the output dataset.

Filter/Group by Array Property:
Group A: Array/Set Name Begins with 'ews'
Group B: Array/Set Name Begins with 'bl'
Group C: Array/Set Name Begins with 'nb'
Group D: Array/Set Name Begins with 'rms'

Fri Oct 10 10:35:40 EDT 2003

63 arrays, 2308 genes in the [input Dataset](#)
230 Genes and 63 arrays passed filters
No genes excluded for being present in less than 95% (60) arrays.
2078 genes excluded where variance is in the lowest 90 percentile (Variance<1.60).

Link to the [output Dataset](#)

1

2

1. Review the subset history on top of the Dataset Display page for the filtering.

2. Click link **History**, a new window will popup with the full dataset history. Review the text.

3. Click **output Dataset** will lead you to the filtered dataset. Close the new window and return to the previous window.

3

Questions:

1. How many genes are filtered out by missing values? How many genes are filtered out by variance?

Lab 5. Hierarchical Clustering

Goal: To cluster genes and/or arrays with the Hierarchical Clustering algorithm.

Filtering/Grouping/Analysis Tools ...

Choose a Tool **Clustering: Hierarchical** and **Proceed**

Interactive Graphical Viewers ...

Choose a Viewer **PCA: Principal Components Analysis** and **View**

Verify that the current dataset is the filtered dataset.
(Gene \geq 95, Variance \geq 90)

1. In the Filtering/Grouping/Analysis section, choose the **Clustering: Hierarchical** Tool.
2. Click on **Proceed**

A new page will be displayed with options for selecting the Similarity/Distance Metric.

Hierarchical Clustering Options ...

Similarity/Distance Metric

Genes: **Correlation (centered - classical Pearson)**

Arrays: **Correlation (centered - classical Pearson)**

Linkage Method: **Average Linkage**

Cluster

3. Choose **Correlation (centered classical Pearson)** to cluster both Genes and Arrays.
4. Click on **Cluster** button.

A new page will be displayed for Hierarchical Clustering progress. When the analysis is done, a **View Clusters** button is displayed on top of the page.

1. Click the **View Clusters** button at the top of the page or the **Click to view result** link at the bottom.

A new page will be displayed with a thumbnail image of the clustering results

1

View Clusters

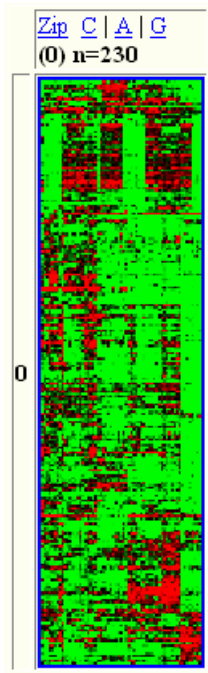
A **View button** should appear above when clustering is finished (a link will also appear at the bottom).

Clustering is performed using a derivative of the **Xcluster** program developed at Stanford University by [Gavin Sherlock](#), Head Microarray Informatics.

Initiating Hierarchical Clustering program...

```
Getting size of data...
Reading Data...
Done reading data...
Assigning Genes to Centroids: iteration 1
Assigning Genes to Centroids: iteration 2
Converged
Making correlations
0
Done Making Correlations
Clustering genes
Done clustering genes
Making correlations
0
Done Making Correlations
Clustering Experiments
10
20
30
40
50
60
Done Clustering Experiments
Outputting cdt file
Done outputting
Finished
```

[Click to view result](#)



1. Click the thumbnail on the page.

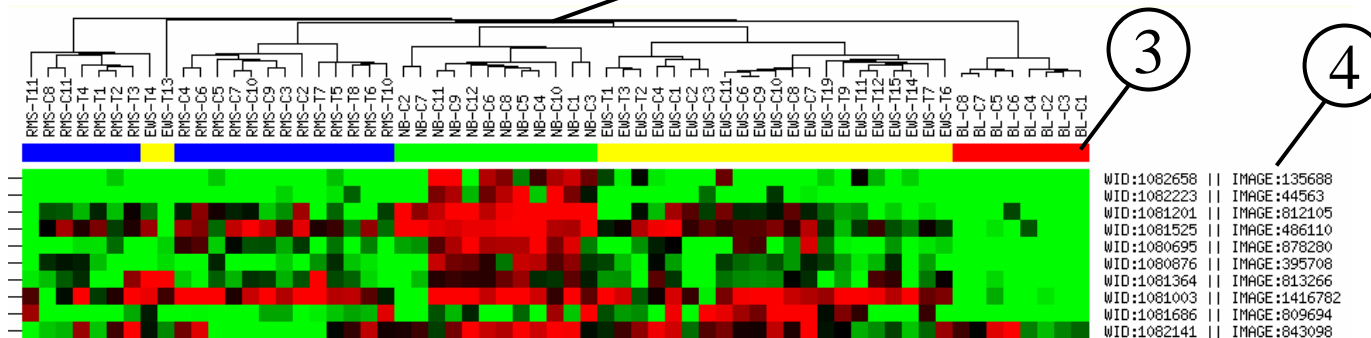
A new browser window will open up to display a enlarged heatmap image, gene trees and array trees.

2. Check the array tree structure. Check the relationship among all the tumor groups.

3. The color bar indicate the grouping information of arrays. Identify the misclassified samples. Speculate possible explanations.

4. Click on the gene annotations on the right. A new window will open up with a feature report page.

5. Close the Feature Report window. Close the Heatmap display window. Return to the thumbnail image window.



Questions:

1. How do the tumor samples cluster together? Can you find duplicate genes that cluster together on the heatmap?

Lab 6. SOM Clustering

Goal: To cluster genes into partitions with 2 dimensional topology using the Self Organizing Map (SOM) algorithm.

Filtering/Grouping/Analysis Tools

Choose a Tool **Clustering: SOM** and **Proceed**

Interactive Graphical Viewers

Choose a Viewer **MDS: MultiDimensional Scaling** and **View**

1. Use the back button of the browser to return to the previous Dataset Display page. Verify that the current dataset is the right dataset. (Gene \geq 95, Variance \geq 90)

2. In the Filtering/Grouping/Analysis section, choose the **Clustering: SOM** Tool

3. Click on **Proceed**

A new page will be displayed with options for SOM.

4. Set X dimension to be 4 and Y dimension to be 3, number of iterations to be 100000. Uncheck the checkbox for Initialize with Randomized Partition.

5. Set the Hierarchical Clustering Options within the SOM clusters. Select **Correlation (centered – classical Pearson)** Metric for Genes and **Not Clustered** for arrays.

6. Click on **Cluster** button.

A new page will be displayed for SOM Clustering progress. When the analysis is done, a **View Clusters** button is displayed on top of the page.

7. Click the **View Clusters** button.

A new page will be displayed with a thumbnail image of the clustering results

Self Organizing Maps Options

Specify X dimension **4**

Specify Y dimension **3**

Number of iterations **100000**

Initialize with Randomized Partition ☐

SOM Elements

Hierarchical Clustering Options

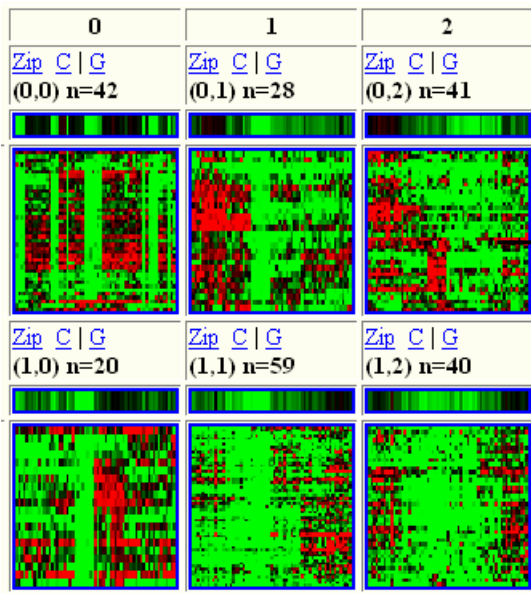
Similarity/Distance Metric

Genes: **Correlation (centered - classical Pearson)**

Arrays: **Not Clustered**

Linkage Method: **Average Linkage**

Cluster



1. Inspect the spatial relationship among the clusters.

2. Click a thumbnail image on the page.

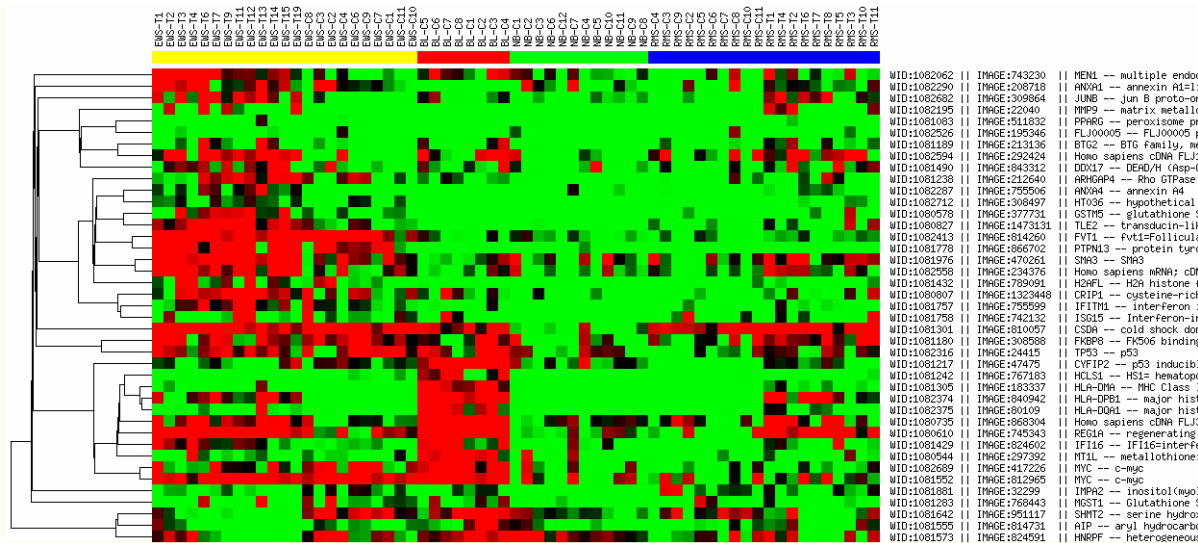
A new browser window will open up to display an enlarged heatmap image and gene tree of the clicked thumbnail image.

3. Look for genes you are familiar with and try to interpret the results.

4. Close the Heatmap display window. Return to the thumbnail image window.

Questions:

1. Do genes in the same partition show a similar expression profile? How are the expression profiles different among different partitions (2-D topology)?




Lab 7. K-means Clustering (Optional)

Goal: To cluster genes into K numbers of partitions using the K-means algorithm.

2

[Expand](#) this Dataset.
Access Datasets in your [Temporary](#) area.

Filtering/Grouping/Analysis Tools 

Choose a Tool and

1. Use the back button of the browser to return to the previous Dataset Display page. Verify that the current dataset is the right dataset. (Gene \geq 95, Variance \geq 90)

2. Click link **Expand this Dataset** above the Filtering/Grouping/Analysis Tools section.

You will then be presented an expanded dataset selection page. You will find the dataset and all the subsets you saved from previous analysis.

3. Click link **Open** open My Grouped Dataset. A mAdb dataset display page will be presented to you. K-means clustering will be performed on the full grouped dataset to show its performance speed advantage.

Label	Origin	Created	Arrays	Genes	
Edit Small, Round Blue Cell Tumors (SRBCTs), Nature Medicine Vol ...	Dataset	Aug 26 6:00:00pm	88	2308	Open
Edit My Grouped Dataset	Subset	Oct 09 5:57:20pm	63	2308	Open
Edit Gene \geq 95%, Variance \geq 90%	Subset	Oct 10 10:35:34 am	63	230	Open

3

Filtering/Grouping/Analysis Tools ...

Choose a Tool **Clustering: Kmeans** and **Proceed**

Interactive Graphical Viewers ...

Choose a Viewer **MDS: MultiDimensional Scaling** and **View**

1. In the Filtering/Grouping/Analysis section, choose the **Clustering: Kmeans** Tool.

2. Click on **Proceed**.

A new page will be displayed with options for Kmeans Clustering.

3. Specify **Number of Nodes** to be 12.

4. Set **Maximum Number of iterations** to be 100.

5. Set the Hierarchical Clustering Options within Kmeans Nodes. Select **Correlation (centered – classical Pearson)** for Genes and **Not Clustered** for arrays.

6. Click on **Cluster** button.

A new page will be displayed for Kmeans Clustering progress. When the analysis is done, a **View Clusters** button is displayed on top of the page.

7. Click the **View Clusters** button.

A new page will be displayed with a thumbnail image of the clustering results.

Kmeans Clustering Options ...

Specify Number of Nodes **12**

Maximum Number of iterations **100**

Kmeans Nodes
Hierarchical Clustering Options ...

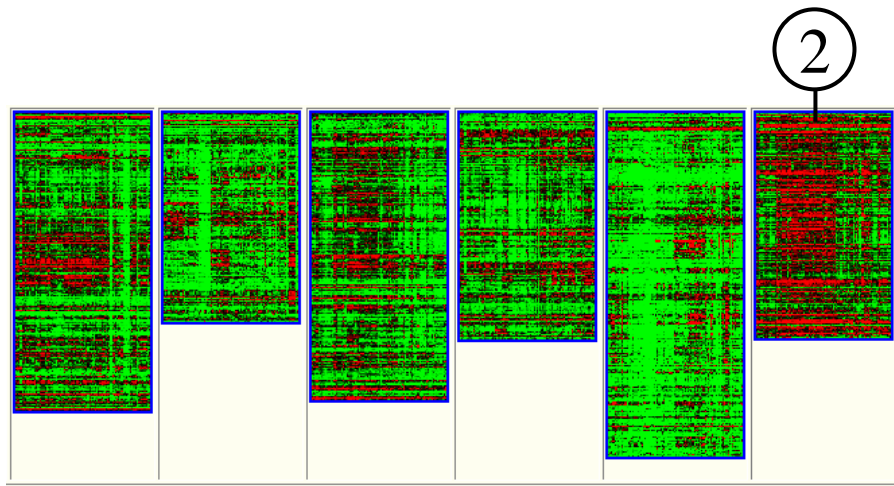
Similarity/Distance Metric

Genes: **Correlation (centered - classical Pearson)**

Arrays: **Not Clustered**

Linkage Method: **Average Linkage**

Cluster



1. Inspect the thumbnail images for the expression patterns within the clusters (Only 6 out of 12 clusters are displayed here).
2. Click thumbnails of interest on the page.

A new browser window will open up to display an enlarged heatmap image and gene tree (not shown here) of the clicked thumbnail image.

3. Find genes you are familiar with in a specific node and try to interpret the results.
4. Close the Heatmap display window. Return to the thumbnail image window.

Questions:

1. Do genes in same partition show a similar expression profile?
2. Are the expression profiles different among different partitions?
3. Can you identify any relationships among partitions?

Lab 8. PCA

Goal: To explore the data structure of the dataset using Principal Component Analysis (PCA).

Filtering/Grouping/Analysis Tools

Choose a Tool: **Additional Filtering Options** and **Proceed**

Interactive Graphical Viewers

Choose a Viewer: **PCA: Principal Components Analysis** and **View**

2 3

Perform PCA on: ☒ Arrays ☐ Genes

Dispersion Matrix: ☐ Correlation ☒ Covariance

Note: Imputing of missing values is not yet available.
Genes with missing values are disgarded from the PCA calculations.

6 **Continue** **Cancel**

PCA was performed on 63 arrays and 230 genes.
No genes contained a missing value.

Proceed to the **3D Viewer**

7

1. Verify that the current dataset is the filtered dataset. (Gene \geq 95, Variance \geq 90)
2. In the **Interactive Graphical Viewers** section, choose the viewer, **PCA: Principal Components Analysis**.

3. Click on **View** button.

A new window, PCA Options, will be displayed with options for the PCA Analysis .

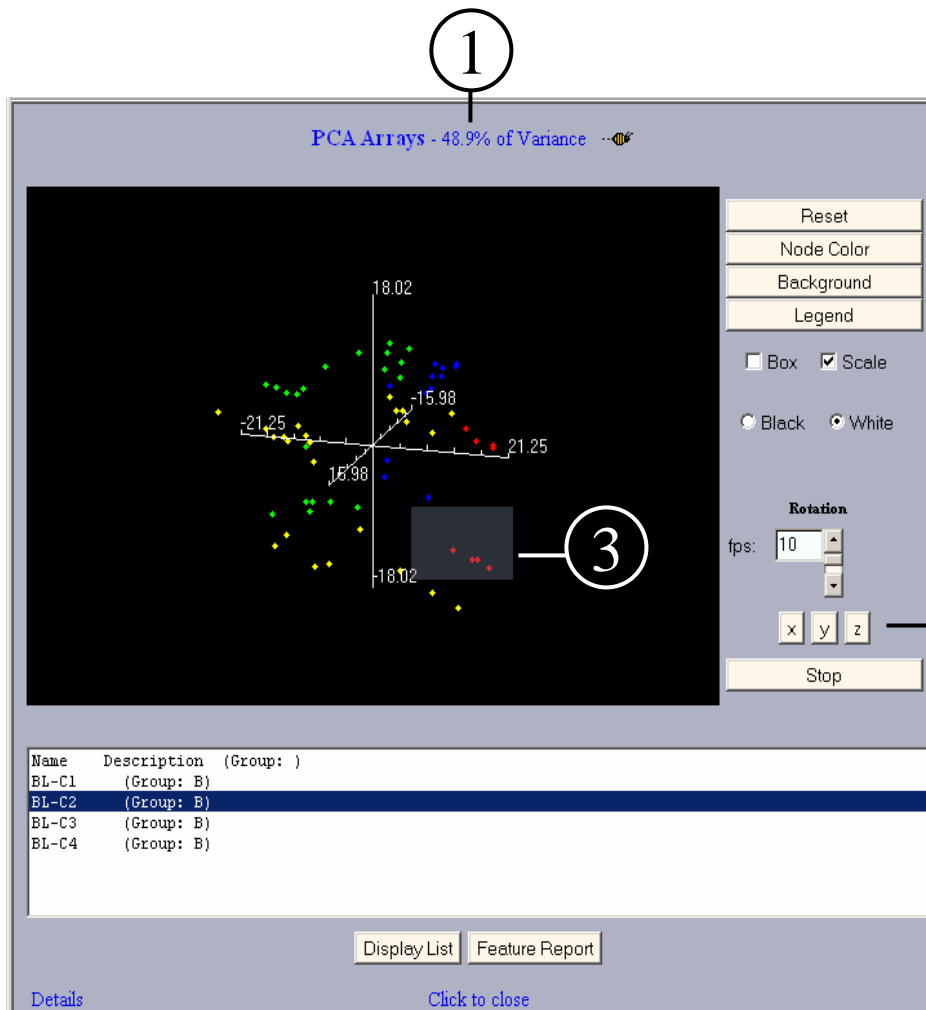
4. Select to perform PCA on **Arrays**.
5. Select Dispersion Matrix of **Covariance**.
6. Click **Continue** button.

A new page, **Waiting for PCA**, will be displayed. When PCA analysis is done, a summary and a new button, **3D Viewer** will be displayed on the page.

7. Click **3D Viewer** button.

Questions:

1. How many genes are used in PCA analysis? (Genes with missing values are not used in PCA)



1. Check the percentage of Variance represented in the 3D plot. Does it capture a large percentage of total variance?

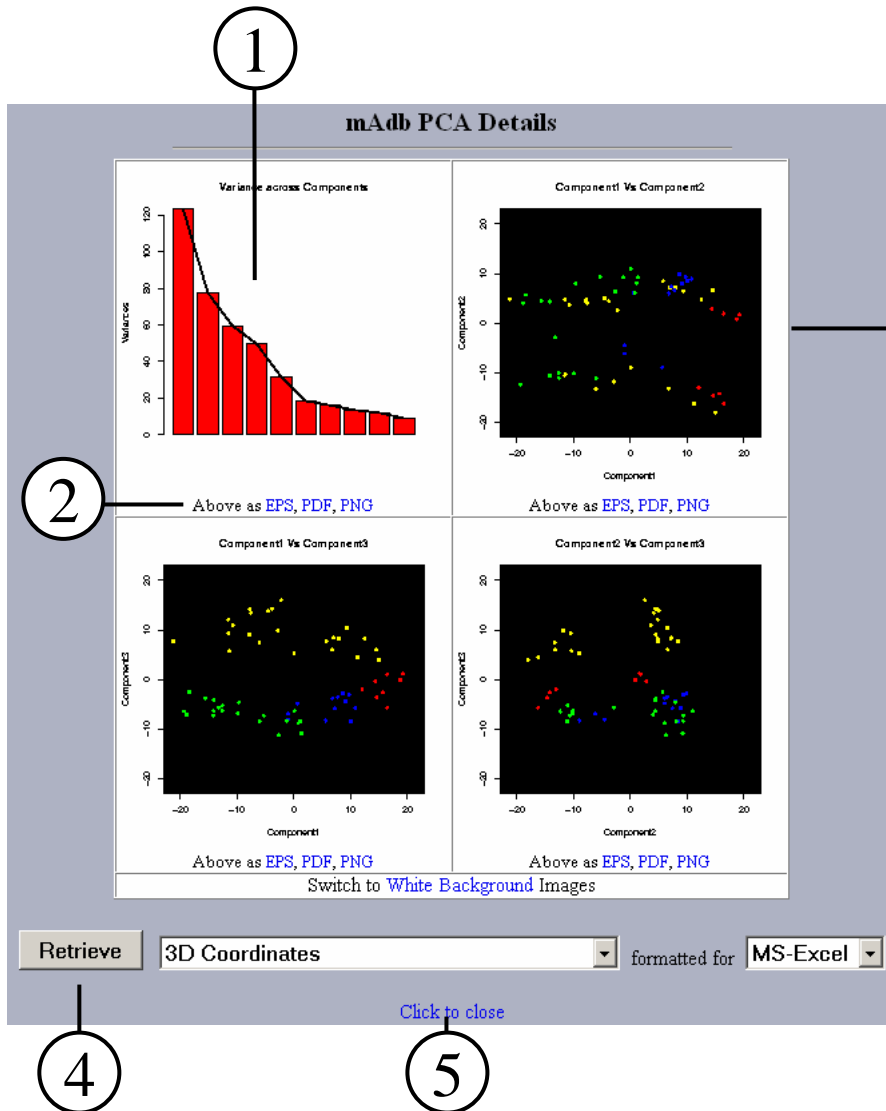
2. Click the **X**, **Y** and **Z** buttons to rotate the 3-D plot. Look for clustering /separation of data. Click **Stop** button.

3. Click and Drag the mouse to highlight an area of the 3 D plot. Data points in the area will be displayed in the text area below the plot.

4. Click the link **Details** on the bottom of the 3D viewer. A new page, **PCA details**, will be displayed with 4 additional plots from PCA analysis. See next page for more description of the PCA details page.

Questions:

1. What is the percentage of variance represented in the first three components?
2. What is the color-coding for each group of samples? Can you see a separation of different groups in 3D plot?



1. The Scree Plot displays the Variance for individual components. Click on the plot will display a new page with an enlarged image.

2. The PDF (Portable Document Format) or PNG (Portable Network Graphics) links under each figure can be used to display or save a larger image of the figure. You can also save a larger image as Encapsulated PostScript using the EPS link

3. The other three plots shown are 2-D plots for each combination of the first 3 components.

4. The **Retrieve** button will retrieve the data back to your local computer. Several options are available. We do not need to retrieve data for this Lab.

5. Click the link **Click to Close** to close the viewer. This will allow you to go back to the starting dataset display page.

Questions:

1. In the scree plot, identify where the slope of variance flattens out (the scree point).

Lab 9. Performing an ANOVA analysis

Goal: To identify differentially expressed genes using class comparison statistical tools.

Choose one or more Projects, select a Tool and Continue or access previously extracted data located in **training01**'s: [Temporary](#) area

1

1. On the mAdb Gateway Page, Click on **Temporary** area to open a list of your Datasets stored in this area.

2. Click on the **Expand** for the “Small Round Blue Cell Tumors (SRBCTs)...” (or, if you are using the other dataset, Expand for the “NEJM – 3 Classes”) to open the list of Subsets for this Dataset.

Temporary Datasets		Created	Containing		Need Help?		Gene Information	
			Arrays	Genes			Refreshed	
Edit	Small, Round Blue Cell Tumors (SRBCTs), Nature Medicine Vol ...	Aug 26 6:00:00pm	88	2308	Open	Expand (1)	Refresh	Aug 26 6:00:00pm
Edit	NEJM - 3 Classes	Aug 26 5:23:18pm	60	1629	Open	Expand (1)	Refresh	Aug 26 5:23:18pm

2

3. Click on the **Open** for the “My Grouped Dataset” subset.

Label		Origin	Created	Containing		Need Help?	
				Arrays	Genes		
Edit	Small, Round Blue Cell Tumors (SRBCTs), Nature Medicine Vol ...	Dataset	Aug 26 6:00:00pm	88	2308	Open	
Edit	My Grouped Dataset	Subset	Oct 16 2:12:33pm	63	2308	Open	History

3

Filtering/Grouping/Analysis Tools

Choose a Tool: **Two or more Group Comparison** and **Proceed**

Choose a Viewer: **MDS: MultiDimensional Scaling** and **View**

1. In the Filtering/Grouping/Analysis section, choose the **Two or more Group Comparison** Tool

2. Click on **Proceed**

A new page will be displayed with options for the statistical comparison analysis. Since this dataset has more than two groups, only the Multiple Group Comparison options for more than two groups will be available for selection.

Statistical Comparison Analysis

Multiple Group Comparison: **One way ANOVA**

Dataset Properties

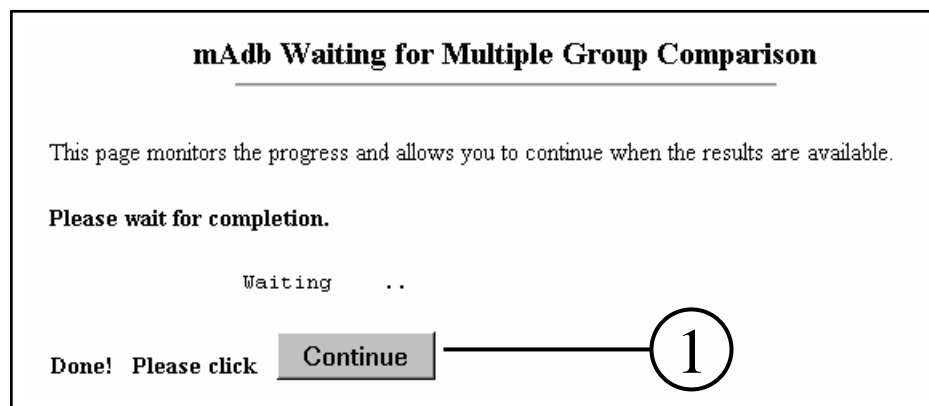
Subset Label: **ANOVA Results**

Proceed

3. Select **One way ANOVA**

4. Analysis results are stored as a new subset. Enter an appropriate Label for this subset.

5. Click on **Proceed**.



A “Waiting” page is displayed while the analysis is being performed. When the analysis is completed, the continue button is displayed.

1. Click on **Continue**. A mAdb Dataset Display page, displaying the newly created subset which contains the ANOVA analysis results will appear.

The three columns, p-Value, Difference and Groups display results from this analysis. The p-Value is the One way ANOVA calculation. The Difference displays the largest difference between group means--this calculation is independent of the ANOVA calculation. The Groups identifies the two groups having this largest mean difference. The default order of the data is from smallest to largest p-Value.

Note that “Show Data Values” has been unchecked for the display shown here.

☒ p-Value

☒ Difference

☒ Groups

Save a Feature Property List (used with the Feature Properties Filtering tool).

→

Records 1 to 25 of 2308 total records displayed.

↓

↑

↓

↑

↓

↑

↓

↑

↓

↑

↓

↑

↓

↑

p-Value	Difference	Groups	Well ID	Feature ID	Gene	Description
9.6276e-22	4.11	A-B	1081848	IMAGE:770394	FCGRT	Fc fragment of IgG, receptor, tra
3.488e-20	2.99	C-D	1082414	IMAGE:784224	FGFR4	FGFR4=Fibroblast growth factor
2.5008e-19	3.59	A-B	1080705	IMAGE:377461	CAV1	caveolin 1, caveolae protein, 22k
2.5733e-18	2.59	A-C	1082413	IMAGE:814260	FVT1	fvt1=Follicular lymphoma variant
1.4459e-17	2.76	C-A	1081462	IMAGE:796258	SGCA	sarcoglycan, alpha (50kDa dystro
5.7703e-17	2.89	A-B	1081004	IMAGE:1435862	MIC2	antigen identified by monoclonal :
8.728e-17	3.14	C-B	1081653	IMAGE:859359	PIG3	quinone oxidoreductase homolog:
1.3957e-16	3.95	D-A	1082509	IMAGE:295985		clone IMAGE:4538214=FLJ2065
4.1114e-16	4.03	A-B	1080566	IMAGE:365826	GAS1	Growth arrest-specific 1

Filtering/Grouping/Analysis Tools

Choose a Tool **Statistics Results Filtering** and **Proceed**

Interactive Graphical Viewers

Choose a Viewer **MDS: MultiDimensional Scaling** and **View**

1. In the Filtering/Grouping/Analysis section, choose the **Statistical Results Filtering Tool**

2. Click on **Proceed**

A new page will be displayed with the options for filtering the statistical results.

Check boxes on the left to activate specific filters

☒ One Way ANOVA p-value **<=** **1e-7**

☐ Group mean Difference **>=** **0**

Subset Label: **Filtered ANOVA p<= 1e-7**
(Optional)

Filter **Cancel**

3. **Check** the box to the left of **One way ANOVA p-value**, select “<=“ and enter the p-value as **0.000001** or **1e-7**.

4. The filtered results will be stored as a new subset. Enter an appropriate Label for this subset.

5. Click on **Filter**. Filtering will be performed and the results stored as a new subset. There is no intermediate “Waiting” page, the new subset will be directly displayed when the Filtering process is completed.

Questions:

1. How many genes are there in the filtered dataset?

In order to facilitate later comparison/filtering of these results with other results, we will save this result as a Feature Property List.

1. Click on Save a Feature Property List.

A new page will be displayed with the options for the Saving a Feature Property List..

Feature Property List

Save a List of: **mAdb Well IDs**

Store the List as: **Global (Available in all Datasets)**

List Label: **SRBCT ANOVA $p \leq 1e-7$**

☐ Overwrite any existing list with the same label

Save

☒ p-Value ☒ Difference

☒ Groups

Save a Feature Property List (used with the Feature Properties Filtering tool).

Records 1 to 25 of 388 total records displayed.

p-Value	Difference	Groups	Well ID	Feature ID	Gene	Description
9.6276e-22	4.11	A-B	1081848	IMAGE:770394	FCGRT	Fc fragment of IgG, receptor, tra
3.488e-20	2.99	C-D	1082414	IMAGE:784224	FGFR4	FGFR4=Fibroblast growth factor
2.5008e-19	3.59	A-B	1080705	IMAGE:377461	CAV1	caveolin 1, caveolae protein, 22k
2.5733e-18	2.59	A-C	1082413	IMAGE:814260	FVT1	fv1=Follicular lymphoma variant
1.4459e-17	2.76	C-A	1081462	IMAGE:796258	SGCA	sarcoglycan, alpha (50kDa distr
5.7703e-17	2.89	A-B	1081004	IMAGE:1435862	MIC2	antigen identified by monoclonal
8.728e-17	3.14	C-B	1081653	IMAGE:859359	PIG3	quinone oxidoreductase homolog
1.3957e-16	3.95	D-A	1082509	IMAGE:295985		clone IMAGE:4538214=FLJ206:
4.1114e-16	4.03	A-B	1080566	IMAGE:365826	GAS1	Growth arrest-specific 1

2. Select **mAdb Well IDs**

3. Select **Global (Available in all Datasets)**

4. Enter an appropriate label to identify this List

5. Click on **Save**

Successfully stored the list.

- Action: **Saved New global List**
- Type: **Well ID**
- Labeled: **SRBCT ANOVA $p \leq 1e-7$**
- Containing: **141** unique, non empty elements

To return to the dataset/subset click

Continue

1

A page indicating that the List was successfully stored and summarizing information about the list will be displayed.

1. Click on **Continue**. This will return you back to the Data Display Page.

Lab 10. Using SAM

Goal: To evaluate statistically significant genes and determine the False Discovery Rate (FDR).

Choose one or more Projects, select a Tool and Continue or access previously extracted data located in **training01**'s: [Temporary](#) area

1

1. On the mAdb Gateway Page, Click on **Temporary** area to open a list of your Datasets stored in this area.

2. Click on the **Expand** for the “Small Round Blue Cell Tumors (SRBCTs)...” (or, if you are using the other dataset, Expand for the “NEJM – 3 Classes”) to open the list of Subsets for this Dataset.

Temporary Datasets		Created	Containing		Need Help?		Gene Information	
			Arrays	Genes			Refreshed	
Edit	Small, Round Blue Cell Tumors (SRBCTs), Nature Medicine Vol ...	Aug 26 6:00:00pm	88	2308	Open	Expand (1)	Refresh	Aug 26 6:00:00pm
Edit	NEJM - 3 Classes	Aug 26 5:23:18pm	60	1629	Open	Expand (1)	Refresh	Aug 26 5:23:18pm

2

3. Click on the **Open** for the “My Grouped Dataset” subset.

Label		Origin	Created	Containing		Need Help?	
				Arrays	Genes		
Edit	Small, Round Blue Cell Tumors (SRBCTs), Nature Medicine Vol ...	Dataset	Aug 26 6:00:00pm	88	2308	Open	
Edit	My Grouped Dataset	Subset	Oct 16 2:12:33pm	63	2308	Open	History

3

Filtering/Grouping/Analysis Tools

Choose a Tool **Filter/Group by Array Properties** and **Proceed**

Interactive Graphical Viewers

Choose a Viewer **MDS: MultiDimensional Scaling** and **View**

1. In the Filtering/Grouping/Analysis section, choose the **Filter/Group by Array Properties Tool**

2. Click on **Proceed**

A new page will be displayed with options for assigning arrays into groups by the naming convention of **Array Name** or **Short Description**.

3. For the SRBC dataset, use BL and NB as matching patterns. Select **Array Name** and **Begins with** from the drop down list for each group.

For the NEJM dataset, use GCB and ABC as matching patterns. Select **Array Name** and **Begins with** from the drop down list for each group.

4. The grouped results are stored as a new subset. Enter an appropriate label for this subset.

5. Click on **Submit**. There is no “Waiting” page, the new grouped subset will be directly displayed when the Group/Filtering process is completed.

Group A **Array Name** **Begins with** BL

Group B **Array Name** **Begins with** NB

Group C **Array Name** **Contains**

Group D **Array Name** **Contains**

Group E **Array Name** **Contains**

Expand the number of possible Group Designations to 10, 15, 20 or 26 groups.

Subset Label: **Two groups for SAM - BL and NB only**

Submit **Cancel**

Filtering/Grouping/Analysis Tools

Choose a Tool **BETA SAM: Significance Analysis of MicroArrays** and **Proceed**

Interactive Graphical Viewers

Choose a Viewer **MDS: MultiDimensional Scaling** and **View**

mAdb SAM Options

SAM help

Warning!!!

Genes with missing values will be automatically removed for SAM analysis. If you want to keep those genes, please use the *"Missing Value Imputation"* tool to first impute the missing data and save a new dataset before calling SAM.

Number of permutations: 100

Set the random seed so that the permutation is reproducible Yes

Continue **Cancel**

1. In the Filtering/Grouping/Analysis section, choose the **SAM: Statistical Analysis for Microarrays** Tool.

2. Click on **Proceed**.

3. Select SAM options for 100 permutations and set the random seed is reproducible.

4. Click on **Continue**.

.

mAdb: Waiting for SAM

This page monitors the progress and allows you to continue when the results are available.

Please wait for completion.

Waiting

SAM Step 1: FDR Calculations Completed!

**Two Class SAM Analysis performed on 20 arrays and 2308 genes.
No genes contained missing values.**

Proceed to the [SAM Step 2](#)

1

SAM Analysis is initiated and A “waiting” page is displayed. When the Analysis is complete, an analysis summary and a button to continue to the next step appears on the page.

1. Click on **SAM Step2**.

Questions:



















1. How many genes contain missing values?

The SAM results are displayed as a table and three graphs. The table shows the number of significant genes, the number of false genes and the false discovery rate (FDR) for each Delta. You can create a subset containing the genes corresponding to one of the models by either clicking on a Shrinkage Delta value or entering a Delta value in the text box and clicking the “Create Subset” button.

The top graph plots the observed $d(i)$ vs. expected $d(i)$ error.

The middle graph plots the FDR vs. the Delta.

The lower graph plots number of significant genes vs. the Delta.

Delta		# of Sig. Genes	# of False Genes	FDR*
0.1		1939	1575.59	0.4300
0.2		1656	1013.94	0.3240
0.3		1334	563.22	0.2234
0.4		1078	303.84	0.1491
0.5		892	159.64	0.0947
0.6		732	81.39	0.0588
0.7		597	41.21	0.0365
0.8		477	19.89	0.0221
0.9		395	10.74	0.0144
1.0		341	6.20	0.0096
1.1		303	3.47	0.0061
1.2		259	2.01	0.0041
1.3		213	1.03	0.0026
1.4		187	0.61	0.0017
1.5		171	0.36	0.0011
1.6		153	0.20	0.0007
1.7		126	0.09	0.0004
1.8		110	0.05	0.0002

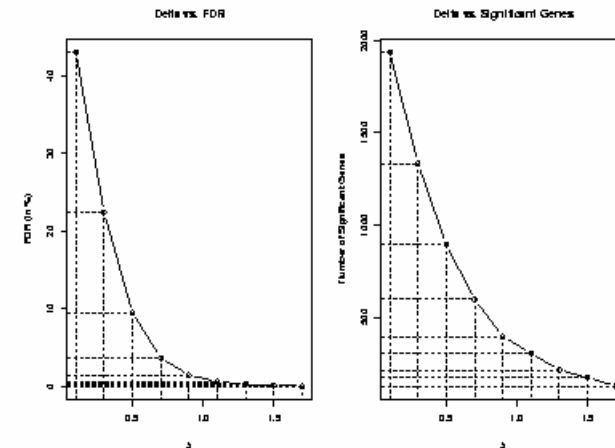
FDR: (Prior Probability) x (# of False Genes)/(# of Sig. Genes)
Prior Probability: 0.5292

The top left graph plots the Delta vs. the FDR

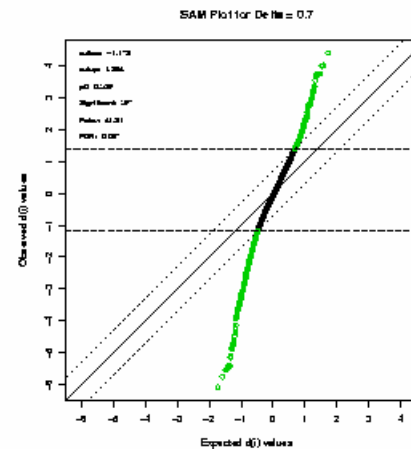
The top right graph plots the Delta vs. the number of significant genes

The lower graph plots number of significant genes vs. the Delta.

The lower graph plots the observed $d(i)$ vs. expected $d(i)$ error.



Above as [EPS](#), [PDF](#), [PNG](#)



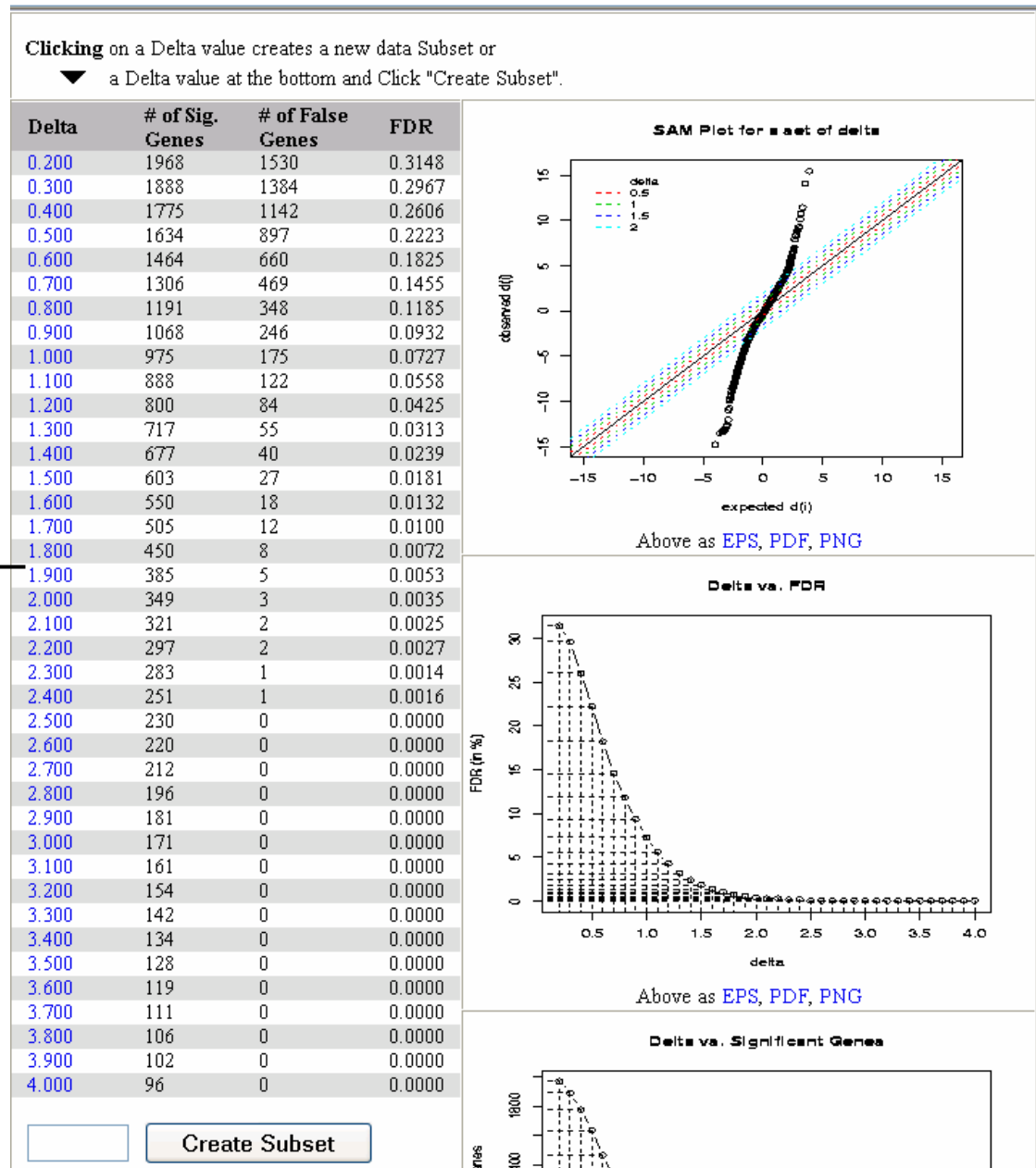
Above as [EPS](#), [PDF](#), [PNG](#)

1. Click on a Delta having a low FDR.

1

Questions:

1. How many genes do you have in the Delta ?
2. What is the FDR for the Delta?



A mAdb Dataset Display page, displaying the newly created SAM subset appears.

Note that **Show Data Values** has been unchecked and the **Background Color** has been set to None for the display shown here.

☐ Show Average(Log2 Ratio)
 ☒ Show Max(Log2 Ratio)-Min(Log2 Ratio)

☐ Show Variance
 ☒ d(i)
 ☒ s(i)

☒ q-value
 ☒ R.difference

[Save](#) a Feature Property List (used with the Feature Properties Filtering tool).

→ Records 1 to 25 of 385 total records displayed.

	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑
Mx-Mn	d(i)	s(i)	q-value	R.difference	Well ID	Feature ID	Map	UniGene								
5.178	-14.7448	0.2402	0	-3.5411	1081310	IMAGE.563673	5q31	Hs.74294								
5.878	-13.4646	0.2883	0	-3.8812	1081201	IMAGE.812105	1q21	Hs.75823								
3.430	-13.3422	0.1698	0	-2.2662	1082036	IMAGE.813742	16p12.1-p11.2	Hs.70500								
4.925	-13.1923	0.2487	0	-3.2815	1081326	IMAGE.784593	2q23.3	Hs.6838								
5.083	-13.0654	0.2866	0	-3.7447	1081525	IMAGE.486110	3q25.1-q25.2	Hs.91747								
5.141	-12.7770	0.2367	0	-3.0241	1082121	IMAGE.377048	2q12-q34	Hs.121576								
6.248	-12.6932	0.3463	0	-4.3960	1082060	IMAGE.629896	5q13	Hs.103042								
4.786	-12.6130	0.2493	0	-3.1439	1081886	IMAGE.504791	6p12.1	Hs.169907								
4.454	-12.0208	0.2422	0	-2.9112	1082481	IMAGE.204545	2p13.1	Hs.8966								

Lab 11. Using PAM

Goal: To evaluate shrunken centroid prediction models and identify sets of genes that best classify sample types.

Choose one or more Projects, select a Tool and Continue or access previously extracted data located in **training01**'s: [Temporary](#) area

1

1. On the mAdb Gateway Page, Click on **Temporary** area to open a list of your Datasets stored in this area.

2. Click on the **Expand** for the “Small Round Blue Cell Tumors (SRBCTs)...” (or, if you are using the other dataset, Expand for the “NEJM – 3 Classes”) to open the list of Subsets for this Dataset.

Temporary Datasets		Created	Containing		Need Help?		Gene Information	
			Arrays	Genes			Refreshed	
Edit	Small, Round Blue Cell Tumors (SRBCTs), Nature Medicine Vol ...	Aug 26 6:00:00pm	88	2308	Open	Expand (1)	Refresh	Aug 26 6:00:00pm
Edit	NEJM - 3 Classes	Aug 26 5:23:18pm	60	1629	Open	Expand (1)	Refresh	Aug 26 5:23:18pm

2

3. Click on the **Open** for the “My Grouped Dataset” subset.

Label		Origin	Created	Containing		Need Help?	
				Arrays	Genes		
Edit	Small, Round Blue Cell Tumors (SRBCTs), Nature Medicine Vol ...	Dataset	Aug 26 6:00:00pm	88	2308	Open	
Edit	My Grouped Dataset	Subset	Oct 16 2:12:33pm	63	2308	Open	History

3

Filtering/Grouping/Analysis Tools

Choose a Tool **PAM: Prediction Analysis for Microarrays** and **Proceed**

Interactive Graphical Viewers

Choose a Viewer **MDS: MultiDimensional Scaling** and **View**

1. In the Filtering/Grouping/Analysis section, choose the **PAM: Prediction Analysis for Microarrays** Tool.

2. Click on **Proceed**

PAM Analysis is initiated and A “waiting” page is displayed. When the Analysis is complete, an analysis summary and a button to continue to the next step appears on the page.

3. Click on **PAM Step2**.

mAdb: Waiting for PAM

This page monitors the progress and allows you to continue when the results are available.

Please wait for completion.

Waiting

PAM Step 1: Training/Cross Validation Done!

PAM 8 Fold Training and Cross Validation was performed on 63 arrays and 2308 genes. No Genes contained missing values, no values were imputed.

Proceed to the **PAM Step 2**

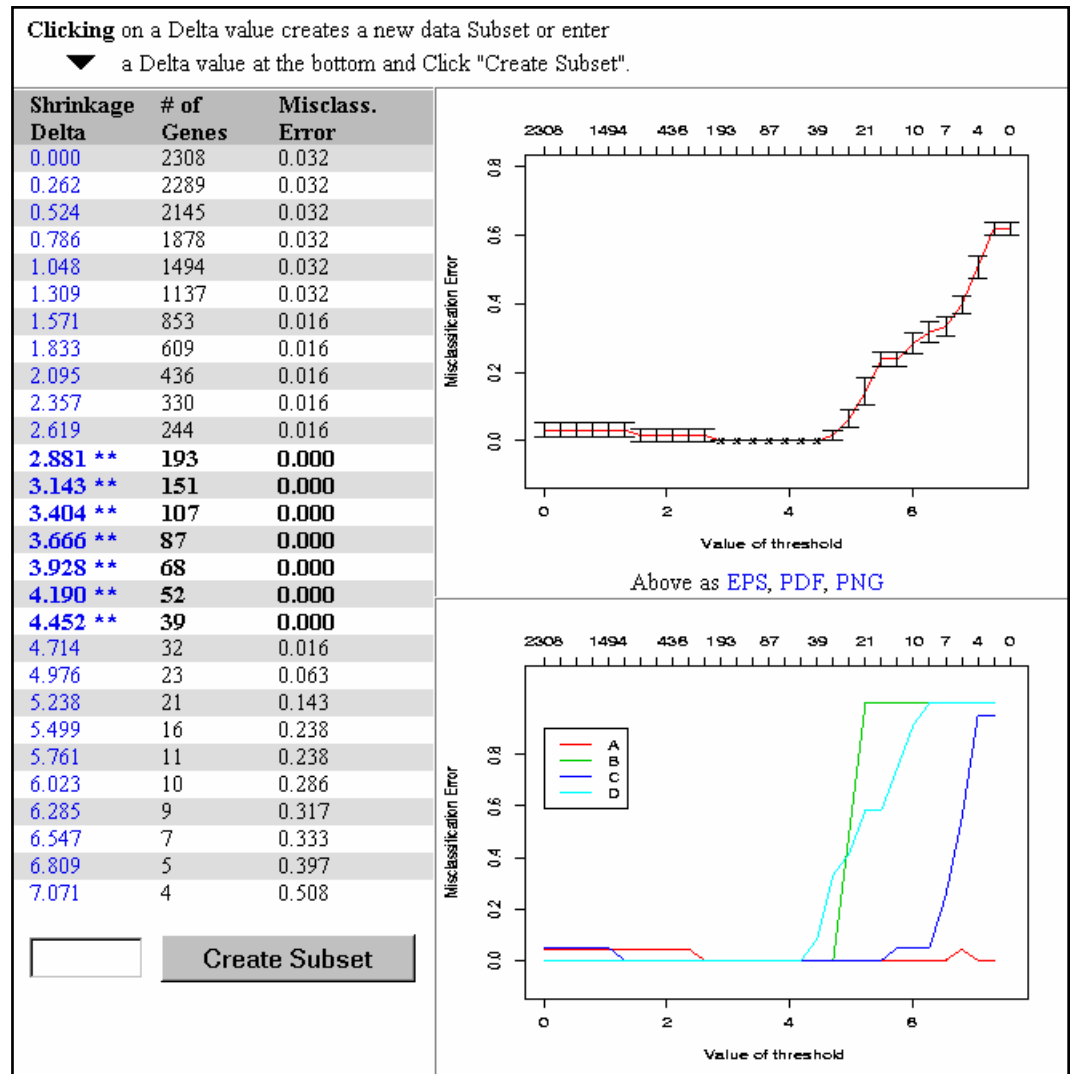
Questions:

1. How many fold of Training and Cross Validation was performed?
2. How many genes contain missing values? How many missing values are imputed for the dataset?

The PAM results are displayed as a table and two graphs. The table shows the Shrinkage Delta (** indicates those having minimum misclassification error), number of genes in the model and the misclassification error based on the K-fold cross validation. You can create a subset containing the genes corresponding to one of the models by either clicking on a Shrinkage Delta value or entering a Delta value in the text box and clicking the “Create Subset” button.

The top graph plots the misclassification error (with error bars) versus the Shrinkage Delta (bottom axis) and the number of Genes (top axis).

The lower graph plots the misclassification error for each group versus the Shrinkage Delta (bottom axis) and the number of Genes (top axis).

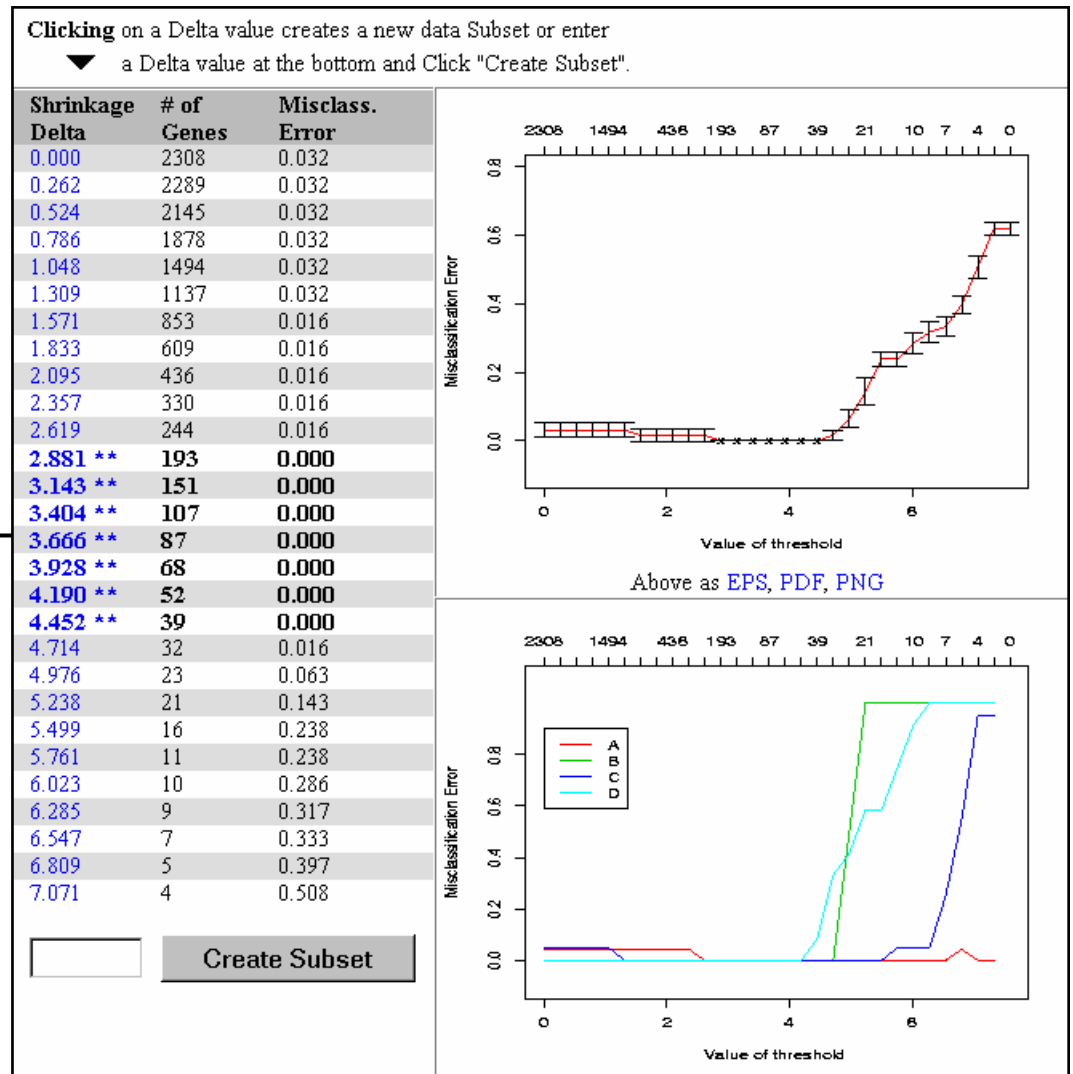


1. Click on a Shrinkage Delta having a minimum misclassification error.

1

Questions:

1. How many genes do you have in the model ?
2. What is the Misclassification Error percentage for the model?



A mAdb Dataset Display page, displaying the newly created PAM subset appears.

The columns A Score, B Score, ... contain the shrunken differences for each group. Non zero values can be used to infer which group or groups a gene's expression value distinguishes.

Note that **Show Data Values** has been unchecked and the **Background Color** has been set to None for the display shown here.

<input checked="" type="checkbox"/> A Score		<input checked="" type="checkbox"/> B Score	
<input checked="" type="checkbox"/> C Score		<input checked="" type="checkbox"/> D Score	

[Save](#) a Feature Property List (used with the Feature Properties Filtering tool).

→ Records 1 to 25 of 87 total records displayed.

↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑
A Score	B Score	C Score	D Score	Well ID	Feature ID	Gene	Description								
0.6527	-0.1732	0.0000	0.0000	1081848	IMAGE:770394	FCGRT	Fc fragment of IgG, receptor, transp								
-0.0429	0.0000	0.6346	0.0000	1082414	IMAGE:784224	FGFR4	FGFR4=Fibroblast growth factor rec								
-0.1131	0.0000	0.6248	0.0000	1080646	IMAGE:296448	IGF2	Insulin-like growth factor 2 (somatom								
0.0000	-0.6186	0.0000	0.0000	1082657	IMAGE:212542		Homo sapiens mRNA; cDNA DKF								
-0.5856	0.0000	0.0000	0.0000	1082509	IMAGE:295985		clone IMAGE:4538214=FLJ20653								
0.5773	0.0000	0.0000	0.0000	1080705	IMAGE:377461	CAV1	caveolin 1, caveolae protein, 22kDa								
0.0000	-0.5739	0.0000	0.0000	1082481	IMAGE:204545	TEM8	tumor endothelial marker 8								
0.0000	-0.5527	0.0000	0.0000	1081310	IMAGE:563673	ALDH7A1	aldehyde dehydrogenase 7 family, m								
0.0000	0.0000	0.5420	0.0000	1080968	IMAGE:207274		Homo sapiens cDNA: FLJ22066 fis								

In order to facilitate later comparison/filtering of these results with other results, we will save this result as a Feature Property List.

☒ A Score
 ☒ B Score
☒ C Score
 ☒ D Score

[Save](#) a Feature Property List (used with the Feature Properties Filtering tool).

Records 1 to 25 of 87 total records displayed.

1. Click on Save a Feature Property List.

1

A Score	B Score	C Score	D Score	Well ID	Feature ID	Gene	Description
0.6527	-0.1732	0.0000	0.0000	1081848	IMAGE:770394	FCGRT	Fc fragment of IgG, receptor, transp
-0.0429	0.0000	0.6346	0.0000	1082414	IMAGE:784224	FGFR4	FGFR4=Fibroblast growth factor rec
-0.1131	0.0000	0.6248	0.0000	1080646	IMAGE:296448	IGF2	Insulin-like growth factor 2 (somato
0.0000	-0.6186	0.0000	0.0000	1082657	IMAGE:212542		Homo sapiens mRNA; cDNA DKF
-0.5856	0.0000	0.0000	0.0000	1082509	IMAGE:295985		clone IMAGE:4538214=FLJ20653
0.5773	0.0000	0.0000	0.0000	1080705	IMAGE:377461	CAV1	caveolin 1, caveolae protein, 22kDa
0.0000	-0.5739	0.0000	0.0000	1082481	IMAGE:204545	TEM8	tumor endothelial marker 8
0.0000	-0.5527	0.0000	0.0000	1081310	IMAGE:563673	ALDH7A1	aldehyde dehydrogenase 7 family, m
0.0000	0.0000	0.5420	0.0000	1080968	IMAGE:207274		Homo sapiens cDNA: FLJ22066 fis

A new page will be displayed with the options for the Saving a Feature Property List..

Feature Property List

Save a List of: **mAddb Well IDs**

Store the List as: **Global (Available in all Datasets)**

List Label: **SRBCT - 87 Gene PAM Model**

☐ Overwrite any existing list with the same label

Save

2. Select **mAddb Well IDs**

3. Select **Global (Available in all Datasets)**

4. Enter an appropriate label to identify this List

5. Click on **Save**

Lab 12. Applying Hierarchical Clustering to the PAM Model


Goal: To use Hierarchical Clustering to explore a PAM Model.

Choose one or more Projects, select a Tool and Continue or access previously extracted data located in **training01's**: [Temporary](#) area

1

1. On the mAdb Gateway Page, Click on **Temporary** area to open a list of your Datasets stored in this area.

2. Click on the **Open** for the “Small Round Blue Cell Tumors (SRBCTs)...” (or, if you are using the other dataset, Expand for the “NEJM – 3 Classes”)

Temporary Datasets		Created	Containing		Need Help? 			Gene Information	
			Arrays	Genes				Refreshed	
Edit	Small, Round Blue Cell Tumors (SRBCTs), Nature Medicine Vol ...	Aug 26 6:00:00pm	88	2308	Open	Expand (1)	Refresh	Aug 26 6:00:00pm	
Edit	NEJM - 3 Classes	Aug 26 5:23:18pm	60	1629	Open	Expand (1)	Refresh	Aug 26 5:23:18pm	

2

Filtering/Grouping/Analysis Tools

Choose a Tool **Feature Property Filtering Options** and **Proceed**

Interactive Graphical Viewers

Choose a Viewer **MDS: MultiDimensional Scaling** and **View**

1. In the Filtering/Grouping/Analysis section, choose the **Feature Property Filtering Options** Tool

2. Click on **Proceed**

A new page will be displayed with options for the Feature Property Filtering.

Check boxes on the left to activate specific filters

☐ **Exclude** Designated Housekeeping Genes

☐ **Include only** Designated Control Features

☐ **Include only** where Well ID = **1080464**

☐ **Include only** where **12345** <= Well ID <= **1040715**

☒ **Include only** where Well ID is in **SRBCT - 87 Gene PAM Model**

*** indicates lists local to this dataset

Subset Label: **87 Gene PAM Model - Complete Dataset**

Filter **Cancel**

3. **Check, Include Only** where Well ID is in Feature List saved in previous Lab (SRBCT – 87 Gene PAM Model for SRBCT dataset).

4. Enter an appropriate Label for the Subset.

5. Click on **Filter**.

Filtering/Grouping/Analysis Tools ...

Choose a Tool **Clustering: Hierarchical** and **Proceed**

Interactive Graphical Viewers ...

Choose a Viewer **PCA: Principal Components Analysis** and **View**

Verify that the current dataset is the right dataset.
(87 Gene PAM Model – Complete Dataset)

1. In the Filtering/Grouping/Analysis section, choose the **Clustering: Hierarchical** Tool
2. Click on **Proceed**

A new page will be displayed with options for selecting the Similarity/Distance Metric.

Hierarchical Clustering Options ...

Similarity/Distance Metric

Genes: **Correlation (centered - classical Pearson)**

Arrays: **Correlation (centered - classical Pearson)**

Linkage Method: **Average Linkage**

Cluster

3. Choose **Correlation (centered classical Pearson)** to cluster both **Genes** and **Arrays**.
4. Click on **Cluster** button.

A new page will be displayed for Hierarchical Clustering progress. When the analysis is done, a **View Clusters** button is displayed on top of the page.

1. Click the **View Clusters** button at the top of the page or the **Click to view result** link at the bottom.

A new page will be displayed with a thumbnail image of the clustering results

View Clusters

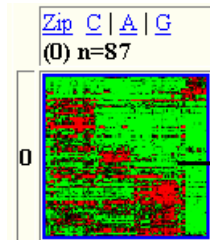
A **View button** should appear above when clustering is finished (a link will also appear at the bottom).

Clustering is performed using a derivative of the **Xcluster** program developed at Stanford University by [Gavin Sherlock](#), Head Microarray Informatics.

Initiating Hierarchical Clustering program...

Getting size of data...
Reading Data...
Done reading data...
Assigning Genes to Centroids: iteration 1
Assigning Genes to Centroids: iteration 2
Converged
Making correlations
0
Done Making Correlations
Clustering genes
Done clustering genes
Making correlations
0
Done Making Correlations
Clustering Experiments
10
20
30
40
50
60
Done Clustering Experiments
Outputting cdt file
Done outputting
Finished

[Click to view result](#)



1

1. Click the thumbnail on the page.

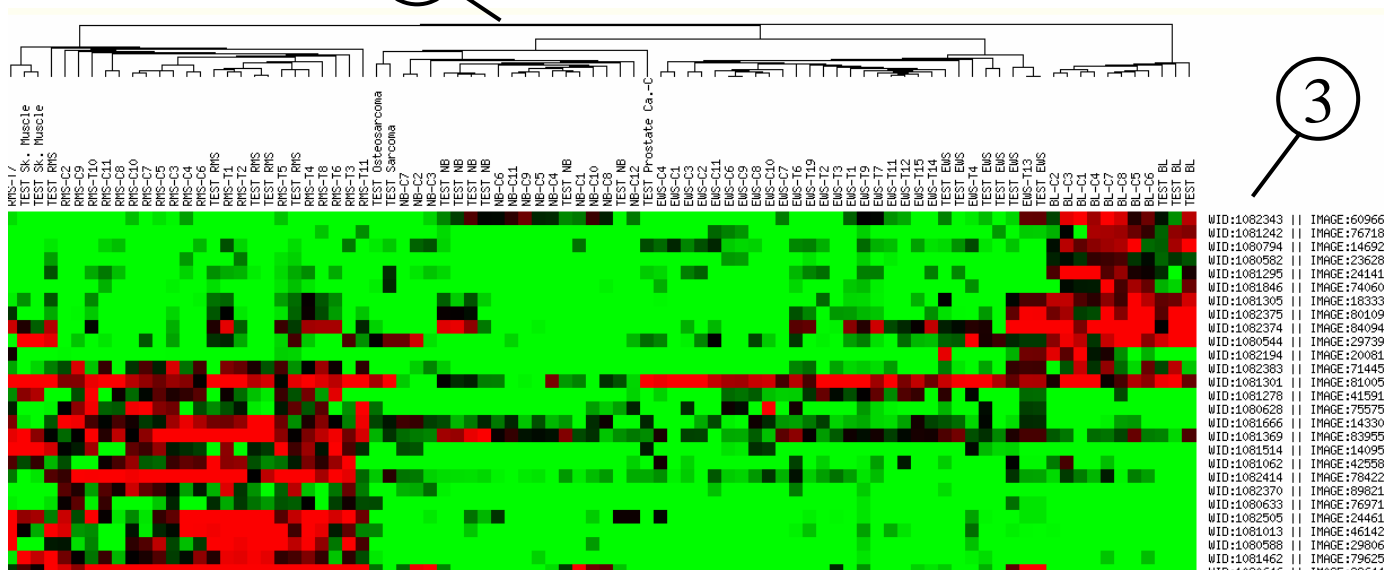
A new browser window will open up to display an enlarged heatmap image and gene tree of the clicked thumbnail image.

2. Check the array tree structure. Check the relationship among all the tumor groups..

3. Click on the gene annotations on the right. A new window will open with a feature report page.

4. Close the Heatmap display window.

2



3

Questions:

1. Review the dendrogram for the samples and identify possible clusters. How does heatmap pattern distinguish the clusters?
2. Review the test arrays not used in the PAM analysis and verify whether they cluster into the right tumor groups.

The SAM results are displayed as a table and three graphs. The table shows the number of significant genes, the number of false genes and the false discovery rate (FDR) for each Delta. You can create a subset containing the genes corresponding to one of the models by either clicking on a Shrinkage Delta value or entering a Delta value in the text box and clicking the “Create Subset” button.

The top graph plots the observed $d(i)$ vs. expected $d(i)$ error.

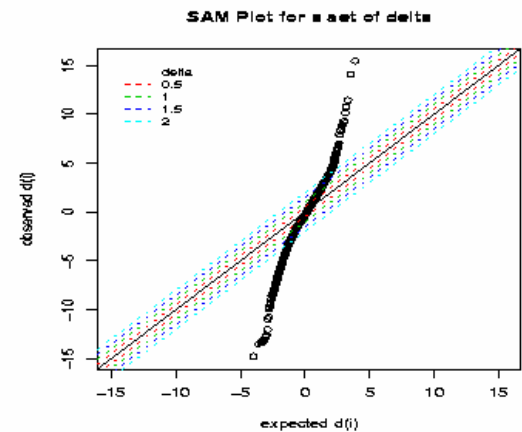
The middle graph plots the FDR vs. the Delta.

The lower graph plots number of significant genes vs. the Delta.

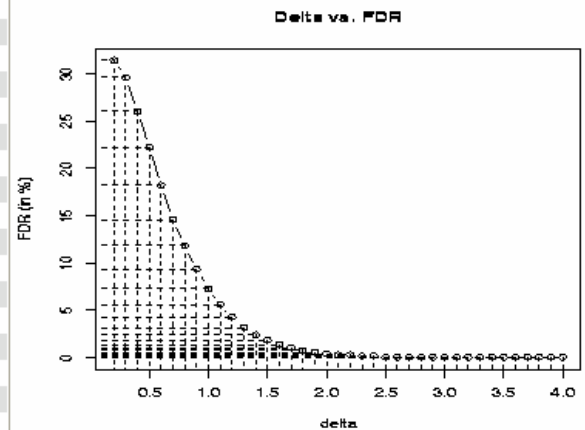
Clicking on a Delta value creates a new data Subset or
▼ a Delta value at the bottom and Click "Create Subset".

Delta	# of Sig. Genes	# of False Genes	FDR
0.200	1968	1530	0.3148
0.300	1888	1384	0.2967
0.400	1775	1142	0.2606
0.500	1634	897	0.2223
0.600	1464	660	0.1825
0.700	1306	469	0.1455
0.800	1191	348	0.1185
0.900	1068	246	0.0932
1.000	975	175	0.0727
1.100	888	122	0.0558
1.200	800	84	0.0425
1.300	717	55	0.0313
1.400	677	40	0.0239
1.500	603	27	0.0181
1.600	550	18	0.0132
1.700	505	12	0.0100
1.800	450	8	0.0072
1.900	385	5	0.0053
2.000	349	3	0.0035
2.100	321	2	0.0025
2.200	297	2	0.0027
2.300	283	1	0.0014
2.400	251	1	0.0016
2.500	230	0	0.0000
2.600	220	0	0.0000
2.700	212	0	0.0000
2.800	196	0	0.0000
2.900	181	0	0.0000
3.000	171	0	0.0000
3.100	161	0	0.0000
3.200	154	0	0.0000
3.300	142	0	0.0000
3.400	134	0	0.0000
3.500	128	0	0.0000
3.600	119	0	0.0000
3.700	111	0	0.0000
3.800	106	0	0.0000
3.900	102	0	0.0000
4.000	96	0	0.0000

Create Subset



Above as [EPS](#), [PDF](#), [PNG](#)



Above as [EPS](#), [PDF](#), [PNG](#)

